EXPLICIT SOLUTION OF RELATIVE ENTROPY WEIGHTED CONTROL

JORIS BIERKENS HILBERT J. KAPPEN
RADBOUD UNIVERSITY NIJMEGEN

ABSTRACT. We expand earlier results by Boué and Dupuis [BD98] where stochastic control problems with a particular cost structure, involving a relative entropy term, are shown to admit a solution by means of a change of measure technique. We provide methods of computing the corresponding optimal control process explicitly. Our results enables us to find solutions for optimal control problems to which the dynamic programming principle can not be applied. The argument is as follows. Minimization of the expectation of a random variable with respect to the underlying probability measure, penalized by relative entropy, may be solved exactly. In the case where the randomness is generated by a standard Brownian motion, this exact solution can be written as a Girsanov density. An explicit expression for the control process may be obtained in terms of the Malliavin derivative of the density process. The theory is applied to the problem of minimizing the maximum of a Brownian motion (penalized by the relative entropy). The link to a linear version of the Hamilton-Jacobi-Bellman equation is made for the case of diffusion processes.

1. INTRODUCTION

In this paper we expand earlier results [BD98] that show how stochastic control problems with a particular cost structure, involving a relative entropy term, admit a purely probabilistic solution, without the necessity of applying the dynamic programming principle. We provide two methods to compute an optimal control in this situation. The first method expresses the optimal control as a Malliavin derivative. This enables us to solve control problems in which the dynamic programming principle fails. The second method transforms the problem of finding an optimal control into a linear PDE.

Essential in our approach are the study of control problems by a change of measure technique [BD98, Dav79, Hau86] and the useful properties of relative entropy [DE97, Section 1.4].

1.1. Background.

A well-known approach in solving optimal control problems is by means of the dynamic programming principle, leading in the stochastic, continuous time case to the HJB equation, a nonlinear PDE [FR75, FS06]. Stochastic optimal control problems with a specific cost structure may be reduced to a linear PDE [Fle82], [FS06, Chapter VI]. Such a linear PDE is obtained by a applying a logarithmic transform to the Hamilton-Jacobi-Bellman (HJB) equation; we will refer to this phenomenon as a linearization of the HJB equation. This observation gained new life in recent years as it was picked up by the physics and artificial intelligence community to obtain Monte Carlo methods for solving stochastic control problems [Kap05].

In this paper we show that this linearizing effect may also be obtained from a purely probabilistic perspective. In fact the linearization is only a special case of a much wider class of probabilistic optimization problems that are regularized by a relative entropy term [BD98]. In the setting of Markov chains the linearizing effect of relative entropy weighted optimization was, to our knowledge, first made in [Tod06]. By limiting arguments the connection with diffusions can be made. In Section 5 we show how this result can be obtained without reference to dynamic programming and without the need for discretization.

Key words and phrases: stochastic optimal control, Itô calculus, Malliavin calculus, diffusions, relative entropy, Kullback-Leibler divergence.

MSC2010 subject classifications: primary 93E20; secondary 60H07, 94A17.

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 270327.

First we provide an alternative way of computing an optimal control by using Malliavin derivatives, as we show in Section 4.

1.2. **Outline.** The main argument is as follows. Let $(\Omega, \mathscr{F}, \mathbb{Q})$ denote a probability space and suppose we are given a random variable C on $(\Omega, \mathbb{Q})$ indicating cost. We may change this probability measure to a new probability measure $\mathbb{P}$ but this operation is 'penalized' by a positive factor $\frac{1}{\beta}$ times the relative entropy $\mathscr{H}(\mathbb{P};\mathbb{Q})$ of the new probability measure with respect to the old probability measure. It is a result of direct computation that the 'optimal' probability measure, i.e. the one that minimizes $\mathbb{E}^{\mathbb{P}} C + \frac{1}{\beta}\mathscr{H}(\mathbb{P};\mathbb{Q})$, has a density proportional to $\exp(-\beta C)$ with respect to $\mathbb{Q}$, as described in Section 2. This result may also be found in [DE97].

If we specialize to the case where all the randomness is generated by a Brownian motion, then an application of Girsanov's theorem shows that a change in probability measure depending on some 'control process' corresponds to a relative entropy equal to quadratic control costs. Furthermore any probability density may be obtained using such a Girsanov type change of measure, which holds in particular for the optimal probability measure with density proportional to $\exp(-\beta C)$ with respect to $\mathbb{Q}$. This result was obtained earlier by Boué and Dupuis [BD98]. This material is explained in detail in Section 3. The focus on this paper is on minimizing cost functionals depending on a Brownian motion. As an exception, in Section 3.4 an example is given of application of the theory to processes with jumps.

An explicit expression of the optimal control process in terms of a Malliavin derivative involving the cost random variable C may be given (Section 4). This argument provides us with a new approach to solve a class of control probems with quadratic control costs. Since the form of the cost random variable is not restricted this method may be applied in cases where dynamic programming (i.e. the HJB equation) fails. For example, it is shown that the maximum of a Brownian motion with drift may be minimized by this method in Section 4.2, resulting in an explicit optimal control policy. This example clearly illustrates the novelty of our approach.

The relation of our approach to classical stochastic optimal control (as in [FR75, KS91]) is explained in Section 5. There it is shown that the solution of the state dependent optimal control problem may be expressed as the solution of a linear PDE. This may be contrasted to the nonlinear HJB PDE that is fundamental in classical stochastic control. As explained above this result was obtained earlier [Fle82, Kap05] but derived in an entirely different way, namely by a logarithmic transformation of the nonlinear HJB equation.

To make the paper self contained we provided some background information on relative entropy (Appendix A.1).

1.3. **Notation.** We denote the Euclidean norm in $\mathbb{R}^n$ by $|\cdot|$. For a matrix $A \in \mathbb{R}^{n\times m}$ we write $||A|| := \sup_{\substack{x\in\mathbb{R}^m \\ |x|=1}} |Ax|$ for the usual matrix norm.

The set of all Borel measurable functions mapping a measurable space D into a Borel space E is denoted by $B(D;E)$, and the set of all bounded Borel measurable functions from D into E is denoted by $B_b(D;E)$. Similarly $C(\mathbb{R}^m;\mathbb{R}^n)$ denotes the space of continuous functions from $\mathbb{R}^m$ into $\mathbb{R}^n$ and $C^{1,2}(I\times\mathbb{R}^m;\mathbb{R}^p)$ denotes the space of continuous functions $f(t,x)$ from $I\times\mathbb{R}^m$ into $\mathbb{R}^n$ that are once differentiable with respect to $t \in I \subset \mathbb{R}$ and twice differentiable with respect to $x \in \mathbb{R}^m$, with all these derivatives being continuous.

If $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space we write $\mathbb{E}^{\mathbb{P}}$ for expectation with respect to the probability measure $\mathbb{P}$.

2. Relative entropy weighted optimization

Let $(\Omega, \mathscr{F}, \mathbb{Q})$ be a probability space. Furthermore suppose a real-valued random variable C, bounded from below, is given, indicating *cost*.

We wish to find a probability measure $\mathbb{P}$ that

(i) is absolutely continuous with respect to $\mathbb{Q}$ (denoted by $\mathbb{P} \ll \mathbb{Q}$),
(ii) minimizes the expected cost $\mathbb{E}^{\mathbb{P}} C$, but
(iii) has minimum deviation from the original probability measure $\mathbb{Q}$. We take the relative entropy

$$\mathscr{H}(\mathbb{P};\mathbb{Q}) = \int_\Omega \ln\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)\, d\mathbb{P} = \mathbb{E}^{\mathbb{P}}\left[\ln\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)\right]$$

as a measure of this deviation (see Appendix A.1).

Note that (i) is a constraint and (ii) and (iii) are conflicting optimization targets. Weighing both the expected cost and the relative entropy, we arrive at the following problem:

Problem 2.1 (Relative entropy weighted optimization)**.** Let $\beta > 0$. Find a probability measure $\mathbb{P} \ll \mathbb{Q}$ such that $\mathbb{P}$ minimizes the functional

$$J(\mathbb{P}) = \mathbb{E}^{\mathbb{P}} C + \frac{1}{\beta}\mathscr{H}(\mathbb{P};\mathbb{Q}) = \mathbb{E}^{\mathbb{P}}\left[C + \frac{1}{\beta}\ln\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)\right]. \tag{1}$$

Let $\mathrm{AC}(\mathbb{Q})$ denote the set of all probability measures $\mathbb{P}$ on $(\Omega,\mathscr{F})$ such that $\mathbb{P} \ll \mathbb{Q}$. Then $\mathrm{AC}(\mathbb{Q})$ is a convex set. The following result, which may also be found in [DE97], says pretty much everything there is to say about this general situation.

Theorem 2.2. *Let $\mathbb{P}^*$ be the measure given by*

$$\frac{d\mathbb{P}^*}{d\mathbb{Q}} = Z^* := \frac{\exp(-\beta C)}{\mathbb{E}^{\mathbb{Q}}\exp(-\beta C)}. \tag{2}$$

Then

(i) *for any $\mathbb{P} \in \mathrm{AC}(\mathbb{Q})$, we have*

$$J(\mathbb{P}) = \frac{1}{\beta}\mathscr{H}(\mathbb{P};\mathbb{P}^*) - \frac{1}{\beta}\ln \mathbb{E}^{\mathbb{Q}}\exp(-\beta C). \tag{3}$$

In particular,

(ii) *J is a strictly convex function over* $\mathrm{AC}(\mathbb{Q})$*,*
(iii) *$\mathbb{P}^*$ solves Problem 2.1, and*
(iv)

$$J(\mathbb{P}^*) = -\frac{1}{\beta}\ln\left(\mathbb{E}^{\mathbb{Q}}\exp(-\beta C)\right). \tag{4}$$

Proof. We prove (i); the other results follow immediately from the properties of relative entropy (Proposition A.5). Write $K = \mathbb{E}^{\mathbb{Q}}\exp(-\beta C)$. To see (3), note that for $d\mathbb{P}/d\mathbb{Q} = Z$,

$$\begin{aligned}\mathscr{H}(\mathbb{P};\mathbb{P}^*) &= \int_\Omega Z\left(\ln Z - \ln Z^*\right)\, d\mathbb{Q} = \mathscr{H}(\mathbb{P};\mathbb{Q}) - \int_\Omega Z \ln Z^*\, d\mathbb{Q}\\ &= \mathscr{H}(\mathbb{P};\mathbb{Q}) + \int_\Omega Z\left(\ln K + \beta C\right)\, d\mathbb{Q} = \mathscr{H}(\mathbb{P};\mathbb{Q}) + \ln K + \beta\mathbb{E}^{\mathbb{Q}} ZC.\end{aligned}$$

Hence

$$J(\mathbb{P}) = \mathbb{E}^{\mathbb{Q}} ZC + \frac{1}{\beta}\mathscr{H}(\mathbb{P};\mathbb{Q}) = \frac{1}{\beta}\mathscr{H}(\mathbb{P};\mathbb{P}^*) - \frac{1}{\beta}\ln K.$$

□

3. Dynamic relative entropy weighted optimization

In this section, we consider the important special situation where all randomness is generated by a multi-dimensional Brownian motion. Changes of measure (satisfying mild conditions) may in this case be expressed as a Girsanov type transformation (see Lemma 3.4 below). The stochastic process appearing in the exponent of the Girsanov density will constitute the 'control process'. Another crucial

observation is that the relative entropy of such a transformation is given by the squared control costs (Proposition 3.8 (i)).

Also we will derive different (but obviously related) explicit expressions for the optimal control process, namely as a time derivative of an expectation (Proposition 3.8(iii)), as a Malliavin derivative (Theorem 4.2) and in the following section as the derivative of a solution to a PDE (Theorem 5.6).

3.1. **Setting.** Let $(\Omega, \mathscr{F}, \mathbb{Q})$ be a probability space. Let $W = (W_t)_{0 \le t \le T}$ define a $\mathbb{Q}$-standard Brownian motion in $\mathbb{R}^p$. Let $(\mathscr{F}_t)_{0 \le t \le T}$ denote the filtration generated by W, and let $\mathscr{F}_\infty := \sigma(\cup_{t \ge 0} \mathscr{F}_t)$.

Let $\mathscr{U}$ denote the set of $\mathbb{R}^p$-valued progressively measurable stochastic processes U such that the process (Z_t^U) defined by

$$Z_t^U := \exp\left(\sum_{i=1}^p \int_0^t U_s^i \, dW_s^i - \tfrac{1}{2} \int_0^t |U_s|^2 \, ds \right), \quad t \ge 0,$$

is a martingale. In particular if U satisfies the *Novikov condition*,

$$\mathbb{E}^x \left[\exp\left(\tfrac{1}{2} \int_0^\infty |U_s|^2 \, ds \right) \right] < \infty,$$

then $U \in \mathscr{U}$ (see [KS91, Proposition 3.5.12]). The set $\mathscr{U}$ will be called the *set of controls* and $U \in \mathscr{U}$ will be a *control process.*

By Girsanov's theorem [KS91, Theorem 3.5.1], there exists a probability measure $\mathbb{P}^U$ defined by the Radon-Nikodým derivative

$$\frac{d\mathbb{P}^U}{d\mathbb{Q}} = Z^U = \exp\left(\sum_{i=1}^p \int_0^\infty U_s^i \, dW_s^i - \tfrac{1}{2} \int_0^\infty |U_s|^2 \, ds \right), \tag{5}$$

with respect to which the process

$$t \mapsto W_t^U := W_t - \int_0^t U_s \, ds, \quad t \ge 0, \tag{6}$$

is a standard Brownian motion. Let $\mathbb{E}^U$ be a shorthand notation for $\mathbb{E}^{\mathbb{P}^U}$.

Suppose a random variable C indicating cost is provided, which is bounded from below and square integrable with respect to $\mathbb{Q}$. Define the *cost function* by

$$J(U) := \mathbb{E}^U C + \frac{1}{\beta} \mathscr{H}(\mathbb{P}^U; \mathbb{Q}) = \mathbb{E}^U \left[C + \frac{1}{2\beta} \int_0^\infty |U_s|^2 \, ds \right], \tag{7}$$

where the equality is a result of Proposition 3.8 (i). We consider the following problem.

Problem 3.1 (Dynamic relative entropy weighted optimization)**.** Find the *optimal value* J^* defined by

$$J^* := \inf_{U \in \mathscr{U}} J(U), \tag{8}$$

and, provided it exists, a unique minimizer $U^* \in \arg\min_{U \in \mathscr{U}} J(U)$.

Note the similarity to Problem 2.1. The main difference between the two problems is that in Problem 3.1, we restrict the possible probability measures to those parametrized by $U \in \mathscr{U}$, through their density given by (5).

3.2. **Main result.** We are now ready to state the main result of this section. First we collect the ingredients.

Hypothesis 3.2.
- (i) Let $(\Omega, \mathscr{F}, \mathbb{Q})$ be a probability space, on which a p-dimensional standard Brownian motion $(W_t)_{t \ge 0}$ is defined;
- (ii) Let $(\mathscr{F}_t)_{t \ge 0}$ be the filtration generated by W;
- (iii) Let C be an $\mathscr{F}_\infty$-measurable random variable which is bounded from below and such that $\mathbb{E}^{\mathbb{Q}}|C| < \infty$;
- (iv) Let $\beta > 0$.

Theorem 3.3. *Suppose the conditions of Hypothesis 3.2 are satisfied. Then there exists a square integrable, $(\mathscr{F}_t)_{t\geq 0}$ adapted stochastic process $U^* \in \mathscr{U}$ that solves Problem 3.1. The corresponding probability measure $\mathbb{P}^* := \mathbb{P}^{U^*}$ is given by* (2) *and the optimal value is given by* (4)*. In particular $\mathbb{P}^*$ is optimal among all probability measures that are equivalent with $\mathbb{Q}$, in the sense that it solves Problem 2.1. If V is another stochastic process solving Problem 3.1 then U^* and V are indistinguishable.*

Before we prove Theorem 3.3 we provide a key lemma that is most helpful in establishing the existence of an optimal U.

Lemma 3.4. *Suppose $(\Omega, \mathscr{F}, \mathbb{Q})$ is a probability space on which a d-dimensional standard Brownian motion $(B_t)_{t\geq 0}$ is defined. Let $(\mathscr{F}_t)_{t\geq 0}$ be the complete filtration (i.e. including all null-sets) generated by $(B_t)_{t\geq 0}$.*

Suppose a nonnegative random variable Z is given such that

(i) $\mathbb{E}^{\mathbb{Q}}|\ln Z| < \infty$;
(ii) *Z is $\mathscr{F}_\infty$-measurable.*
(iii) $\mathbb{E}^{\mathbb{Q}} Z = 1$.

Let $\mathbb{P}$ be a probability measure with density Z with respect to $\mathbb{Q}$, so $\frac{d\mathbb{P}}{d\mathbb{Q}} = Z$, and define the conditional density process (Z_t) by $Z_t := \mathbb{E}^{\mathbb{Q}}[Z|\mathscr{F}_t]$, $t \geq 0$. Then Z_t is $\mathbb{Q}$-a.s. continuous. Furthermore there exists a unique $\mathbb{R}^d$-valued, progressively measurable stochastic process θ such that

(a) *the following expression holds:*

$$Z_t = \exp\left(\sum_{i=1}^d \int_0^t \theta_s^i \, dB_s^i - \tfrac{1}{2}\int_0^t \sum_{i=1}^d |\theta_s|^2 \, ds\right), \quad \text{for all } t \geq 0,\ \mathbb{Q}\text{-almost surely.} \tag{9}$$

(b) *the process $\widetilde{B}_t = B_t - \int_0^t \theta_s \, ds$, $t \geq 0$, is a $\mathbb{P}$-Brownian motion.*
(c) *θ is square integrable, and $\mathbb{E}^{\mathbb{Q}}\left[\frac{1}{2}\int_0^\infty |\theta_s|^2 \, ds\right] = -\mathbb{E}^{\mathbb{Q}} \ln Z$.*

Proof. Define a uniformly integrable martingale (Z_t) by $Z_t := \mathbb{E}^{\mathbb{Q}}[Z|\mathscr{F}_t]$. By the martingale representation theorem for Brownian martingales [Kal02, Theorem 18.10], (Z_t) is $\mathbb{Q}$-a.s. continuous (and therefore progressively measurable) and there exists a unique $\mathbb{R}^d$-valued, progressively measurable process ϕ satisfying $\int_0^\infty |\phi_s|^2 \, ds < \infty$, $\mathbb{Q}$-a.s. such that

$$Z_t = 1 + \sum_{i=1}^d \int_0^t \phi_s^i \, dB_s^i, \quad t \geq 0, \quad \mathbb{Q}\text{-almost surely.} \tag{10}$$

The condition $\mathbb{E}^{\mathbb{Q}}|\ln Z| < \infty$ implies that $Z > 0$, $\mathbb{Q}$-almost surely. By Lemma 3.5, $Z_t > 0$ for all $t \geq 0$, $\mathbb{Q}$-almost surely. Ignoring the $\mathbb{Q}$-null set where $Z_t = 0$, define $\theta_t := \frac{\phi_t}{Z_t}$ for $t \geq 0$. Note that θ is progressively measurable, since ϕ and Z are. With this choice of θ, we have $dZ_t = \sum_{i=1}^d \theta_t^i Z_t \, dB_t^i$, for $t \geq 0$, with solution (9), $\mathbb{Q}$-almost surely. Part (b) is then a direct consequence of Girsanov's theorem (see [Kal02, Theorem 18.19]). Since $\mathbb{E}^{\mathbb{Q}}|\ln Z| < \infty$ we may compute $-\mathbb{E}^{\mathbb{Q}} \ln Z$ as

$$-\mathbb{E}^{\mathbb{Q}} \ln Z = \mathbb{E}^{\mathbb{Q}}\left[\frac{1}{2}\int_0^\infty |\theta_s|^2 \, ds - \sum_{i=1}^d \int_0^\infty \theta_s^i \, dB_s^i\right] = \tfrac{1}{2}\mathbb{E}^{\mathbb{Q}} \int_0^\infty |\theta_s|^2 \, ds.$$

□

In the proof of Lemma 3.4 we used the following lemma.

Lemma 3.5. *Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{Q})$ be a filtered probability space and let Z be a $\mathbb{Q}$-density on Ω, i.e. $Z : \Omega \to \mathbb{R}$, $Z \geq 0$, $\mathbb{Q}$-a.s. and $\mathbb{E}^{\mathbb{Q}} Z = 1$.*

If $Z > 0$, $\mathbb{Q}$-a.s., then $\inf_{t\geq 0} Z_t > 0$, $\mathbb{Q}$-a.s., where $Z_t := \mathbb{E}^{\mathbb{Q}}[Z|\mathscr{F}_t]$.

Proof. Define $\mathbb{P}$ by $\frac{d\mathbb{P}}{d\mathbb{Q}} = Z$. By [Kal02, Lemma 18.17] $\inf_{0\le t\le n} Z_t > 0$, $\mathbb{P}$-a.s. for all $n \ge 0$. Therefore

$$A = \left\{\inf_{t\ge 0} Z_t = 0\right\} = \cup_{n\in\mathbb{N}} \left\{\inf_{0\le t\le n} Z_t = 0\right\}$$

is a $\mathbb{P}$-null set. We have $\mathbb{E}^{\mathbb{Q}}[Z\mathbb{1}_A] = \mathbb{P}(A) = 0$. If $\mathbb{Q}(A) > 0$, then $Z(\omega) = 0$ for $\mathbb{Q}$-almost all $\omega \in A$, which is a contradiction. Hence $\mathbb{Q}(A) = 0$. □

The condition $\mathbb{E}^{\mathbb{Q}}|\ln Z| < \infty$ excludes the situation where $\mathbb{Q}(Z = 0) > 0$. Note in particular if $Z \ge c$, $\mathbb{Q}$-a.s. for some $c > 0$, then by Lemma 3.4, we have $\frac{1}{2}\mathbb{E}^{\mathbb{Q}} \int_0^\infty |\theta_s|^2 \le -\ln c$. To illustrate the condition $\mathbb{E}^{\mathbb{Q}}|\ln Z| < \infty$ we provide an example where this condition is not satisfied, and find that $\mathbb{P}(Z_t = 0) > 0$ and therefore the definition of θ becomes problematic. This should not be surprising, since the expression (9) can not become zero for well-behaved (i.e. square integrable) θ.

Example 3.6 (Where the density process becomes zero). Suppose, for some $a \in \mathbb{R}$ and a standard Brownian motion B,

$$Z = \begin{cases} k & \text{if } B_T \ge a \\ 0 & \text{otherwise.} \end{cases}$$

where k is a normalizing constant and $T > 0$. Note that $\mathbb{Q}(Z = 0) > 0$ and hence $\mathbb{E}^{\mathbb{Q}}|\ln Z| = \infty$. We compute

$$Z_t = \mathbb{E}^{\mathbb{Q}}[Z|\mathscr{F}_t] = k\,\mathbb{Q}(B_{T-t} \ge a - x)|_{x=B_t} = \frac{k}{\sqrt{2\pi}} \int_{\frac{a-B_t}{\sqrt{T-t}}}^\infty \exp\left(-\xi^2/2\right)\, d\xi, \quad 0 \le t < T$$

and

$$Z_t = \mathbb{E}^{\mathbb{Q}}[Z|\mathscr{F}_t] = \begin{cases} 1 & \text{if } B_T \ge a \\ 0 & \text{otherwise.} \end{cases} \quad \text{for } t \ge T.$$

The following lemma will be used to establish uniqueness of the solution.

Lemma 3.7. *Suppose U and V are $\mathbb{R}^p$-valued stochastic processes in $\mathscr{U}$. Then*

$$\mathscr{H}(\mathbb{P}^U;\mathbb{P}^V) = \mathbb{E}^U \int_0^\infty (U_s - V_s)^2\, ds.$$

In particular, if $\mathbb{P}^U = \mathbb{P}^V$ then U and V are indistinguishable.

Proof. We compute, for simplicity in the case $p = 1$,

$$\begin{aligned}
\mathscr{H}(\mathbb{P}^U;\mathbb{P}^V) &= \mathbb{E}^U\left[\ln\left(\frac{d\mathbb{P}^U}{d\mathbb{P}^V}\right)\right] = \mathbb{E}^U\left[\ln\left(\frac{d\mathbb{P}^U}{d\mathbb{Q}}\frac{d\mathbb{Q}}{d\mathbb{P}^V}\right)\right] \\
&= \mathbb{E}^U\left[\int_0^\infty U_s\, dW_s - \tfrac{1}{2}\int_0^\infty U_s^2\, ds - \int_0^\infty V_s\, dW_s + \tfrac{1}{2}\int_0^\infty V_s^2\, ds\right] \\
&= \mathbb{E}^U\left[\int_0^\infty U_s\, dW_s^U + \tfrac{1}{2}\int_0^\infty U_s^2\, ds - \int_0^\infty V_s\, dW_s^U - \int_0^\infty U_s V_s\, ds + \tfrac{1}{2}\int_0^\infty V_s^2\right] \\
&= \mathbb{E}^U \int_0^\infty (U_s - V_s)^2\, ds.
\end{aligned}$$

□

Proof of Theorem 3.3: As already noted, Problem 3.1 is a version of Problem 2.1 but restricted to the set of probability measures with density $\mathbb{P}^U$ for $U \in \mathscr{U}$. So if we can find a square integrable $U \in \mathscr{U}$ for which $\mathbb{P}^U$ has the density function given by (2), i.e. one that solves Problem 2.1 because of Theorem 2.2, then it only remains to show uniqueness of such a U.

Therefore we simply define our candidate density function Z^* by (2). Note that Z^* is $\mathbb{Q}$-square integrable since C is bounded from below. Note that $\mathbb{E}^{\mathbb{Q}}\exp(-\beta C) < \infty$ since C is bounded from below. Since $\mathbb{E}^{\mathbb{Q}}|C| < \infty$ it follows that $\mathbb{Q}(C < \infty) = 1$, so that $\mathbb{E}^{\mathbb{Q}}\exp(-\beta C) > 0$. Hence

$$\mathbb{E}^{\mathbb{Q}}|\ln Z^*| = \mathbb{E}^{\mathbb{Q}}\left|-\beta C - \ln \mathbb{E}^{\mathbb{Q}}\exp(-\beta C)\right| \le \beta\mathbb{E}^{\mathbb{Q}}|C| + \left|\ln \mathbb{E}^{\mathbb{Q}}\exp(-\beta C)\right| < \infty.$$

By Lemma 3.4 there exists a square integrable, adapted, $\mathbb{R}^p$-valued stochastic process U^* such that $Z_t^* := \mathbb{E}^{\mathbb{Q}}[Z^*|\mathscr{F}_t]$ has the form (9). Since by definition (Z_t^*) is a martingale, $U^* \in \mathscr{U}$. By Lemma 3.7, U^* is unique up to indistinguishability. The expression for the optimal value follows directly from Theorem 2.2. □

Here we give some useful equalities, which hold for any $U \in \mathscr{U}$ (so in particular for U^*).

Proposition 3.8. *The following relations hold for any $U \in \mathscr{U}$.*

(i) $\mathscr{H}(\mathbb{P}^U;\mathbb{Q}) = \mathbb{E}^U\left[\frac{1}{2}\int_0^\infty |U_s|^2\, ds\right]$;
(ii) $\mathbb{E}^{\mathbb{Q}}\left[Z^U W_r|\mathscr{F}_t\right] = Z_t^U W_t + \int_t^r \mathbb{E}^{\mathbb{Q}}[U_s Z_s^U|\mathscr{F}_t]\, ds$ *for* $0 \le t \le r$;
(iii) $U_t = \dfrac{\frac{d}{dr}\mathbb{E}^{\mathbb{Q}}[W_r Z^U|\mathscr{F}_t]\Big|_{r=t}}{Z_t^U} = \dfrac{\frac{d}{dr}\mathbb{E}^{\mathbb{Q}}[W_r Z_r^U|\mathscr{F}_t]\Big|_{r=t}}{Z_t^U}$ *a.s. for* $t \ge 0$;

Proof. (i): This is a direct consequence of Lemma 3.7, by taking $V \equiv 0$.

(ii): In the remainder of the proof we fix $U \in \mathscr{U}$ and omit the superscripts U in Z^U etc. Let $t \ge 0$ and $r > t$. Define a stochastic process $(Y_s)_{s\ge t}$ by

$$Y_s = \begin{cases} W_s & t \le s \le r, \\ W_r & s > r \end{cases}$$

and note that Y satisfies the equation $dY_s = \mathbb{1}_{s\le r}\, dW_s$ for $s \ge t$. The process (Z_s) is the stochastic exponential of U_s, so $dZ_s = \sum_{i=1}^p U_s^i Z_s\, dW_s^i$. Then using Itô's formula,

$$d(Y_s Z_s) = (\mathbb{1}_{s\le r} Z_s + U_s Z_s Y_s)\, dW_s + \mathbb{1}_{s\le r} U_s Z_s\, ds,$$

so that

$$\mathbb{E}^{\mathbb{Q}}[Z W_r|\mathscr{F}_t] = \mathbb{E}^{\mathbb{Q}}[Z_\infty Y_\infty|\mathscr{F}_t] = Z_t Y_t + \int_t^\infty \mathbb{1}_{s\le r}\mathbb{E}^{\mathbb{Q}}[U_s Z_s|\mathscr{F}_t]\, ds = Z_t W_t + \int_t^r \mathbb{E}^{\mathbb{Q}}[U_s Z_s|\mathscr{F}_t]\, ds.$$

(iii): Taking the derivative of relation (ii) with respect to r and evaluating at $r = t$,

$$\frac{d}{dr}\mathbb{E}^{\mathbb{Q}}[Z W_r|\mathscr{F}_t]\Big|_{r=t} = \mathbb{E}^{\mathbb{Q}}[U_r Z_r|\mathscr{F}_t]\Big|_{r=t} = U_t Z_t.$$

□

3.3. **Example:** $C = \frac{1}{2}(W_T - x^*)^2$**.** We illustrate the theory developed so far on a simple example. Let $T > 0$. Let $(W_t)_{0\le t\le T}$ denote a standard Brownian motion and consider the process $X_t := W_t$. (After a Girsanov change of measure, X will have become a Brownian motion with drift.) For the cost variable $C := \frac{1}{2}(X_T - x^*)^2$, we have $Z^* \propto \exp(-\beta C) = \exp\left(-\frac{1}{2}\beta(W_T - x^*)^2\right)$. Letting K denote a normalization constant, we compute

$$Z_t^* = \mathbb{E}^{\mathbb{Q}}[Z^*|W_t = x] = \frac{1}{K}\mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\tfrac{1}{2}\beta(W_T - x^*)^2\right)\Big| X_t\right] = \frac{1}{K}\mathbb{E}^{\mathbb{Q}}[\exp(-\tfrac{1}{2}\beta(W_{T-t} + x - x^*)^2)]\Big|_{x=W_t}$$
$$= \frac{1}{K\sqrt{\rho(t)}}\exp\left(-\tfrac{1}{2}\beta(W_t - x^*)^2/\rho(t)\right)$$

where $\rho(t) := 1 + \beta^2(T - t)$, $0 \le t \le T$. In this computation we used the Markov property of W and Lemma A.6 (i).

Using Itô's formula,

$$dZ_t^* = -\frac{\beta(W_t - x^*)}{\rho(t)} Z_t^*\, dW_t = -\frac{\beta(X_t - x^*)}{\rho(t)} Z_t^*\, dW_t,$$

or, equivalently,

$$Z_t^* = \exp\left(\int_0^t U_s^*\, dW_s + \frac{1}{2}\int_0^t (U_s^*)^2\, ds\right), \quad 0 \le t \le T,$$

where $U_t^* = -\frac{\beta(X_t - x^*)}{\rho(t)}$, $0 \le t \le T$, so that $\frac{d\mathbb{P}^{U^*}}{d\mathbb{Q}} = Z^*$.

By Theorem 3.3, the process U^* minimizes

$$J(U) = \mathbb{E}^U \left[\tfrac{1}{2}(X_T - x^*)^2 + \frac{1}{2\beta} \int_0^T U^2(s, X_s)\, ds \right],$$

over $U \in \mathscr{U}$, where X satisfies the SDE $dX_t = dW_t = U_t\, dt + dW_t^U$ with W_t^U a Brownian motion under $\mathbb{P}^U$. This is a first example where the optimal solution is computed without application of the Hamilton-Jacobi-Bellman equation.

3.4. **Processes with jumps.** The procedure to obtain a solution to a relative entropy weighted optimization problem may in principle be repeated for other stochastic processes, such as jump processes. We will illustrate this for an example.

Let N be a Poisson random measure on $\mathbb{R}$ with intensity measure ν satisfying $\int_{\mathbb{R}} \nu(dz) < \infty$. Let the cost variable be given by $C := X_T$, where $X_t := \int_0^t \int_{\mathbb{R}} \gamma(z)\, N(ds, dz)$, for $0 \le t \le T$. Let $\beta > 0$. By Theorem 2.2, the density $Z^* \propto \exp(-\beta C)$ is optimal in the sense that it solves Problem 2.1. We wish to see what effect this change of density has on the dynamics.

We compute

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}[\exp(-\beta C)|\mathscr{F}_t] &= \exp(-\beta X_t) \sum_{n=0}^{\infty} \mathbb{P}\left(N([t,T],\mathbb{R}) = n\right) \prod_{j=1}^{n} \int_{\mathbb{R}} \exp\left(-\beta\gamma(z)\right) \nu(dz)/\nu(\mathbb{R}) \\
&= \exp(-\beta X_t) \sum_{n=0}^{\infty} \frac{\mathrm{e}^{-\nu(\mathbb{R})(T-t)} (\nu(R)(T-t))^n}{n!} \left(\int_{\mathbb{R}} \exp\left(-\beta\gamma(z)\right) \nu(dz)/\nu(\mathbb{R}) \right)^n \\
&= \exp(-\beta X_t - \nu(\mathbb{R})(T-t)) \sum_{n=0}^{\infty} \frac{\left(\int_{\mathbb{R}} \exp\left(-\beta\gamma(z)\right)\nu(dz)(T-t)\right)^n}{n!} \\
&= \exp\left(-\beta \int_0^t \int_{\mathbb{R}} \gamma(z)\, N(ds, dz) + (T-t) \int_{\mathbb{R}} \left\{ \exp(-\beta\gamma(z)) - 1 \right\} \nu(dz) \right).
\end{aligned}$$

For the optimal density process this gives

$$Z_t^* := \mathbb{E}^{\mathbb{Q}}[\exp(-\beta C)|\mathscr{F}_t]/\mathbb{E}^{\mathbb{Q}}[\exp(-\beta C)] = \exp\left(-\beta \int_0^t \int_{\mathbb{R}} \gamma(z)\, N(ds, dz) - t \int_{\mathbb{R}} \left\{ \exp(-\beta\gamma(z)) - 1 \right\} \nu(dz) \right).$$

An application of Girsanov's theorem for jump processes [ØS07, Theorem 1.35] gives that the random measure

$$\widetilde{N}^*(dt, dz) = -\exp(-\beta\gamma(z))\nu(dz)\, dt + N(dt, dz)$$

is the compensated random measure corresponding to $N(dt, dz)$ under the optimal probability measure $\mathbb{P}^*$ (as prescribed by Theorem 3.3). In particular, the intensity measure of $N(dt, dz)$ under $\mathbb{P}^*$ is $\nu^*(dz) := \exp(-\beta\gamma(z))\nu(dz)$. The relative entropy of $\mathbb{P}^*$ with respect to $\mathbb{Q}$ may be computed as

$$\begin{aligned}
\mathscr{H}(\mathbb{P}^*; \mathbb{Q}) = \mathbb{E}^* \ln Z_T &= \mathbb{E}^* \left[-\beta \int_0^T \int_{\mathbb{R}} \gamma(z)\, N(ds, dz) - T \int_{\mathbb{R}} \left\{ \exp(-\beta\gamma(z)) - 1 \right\} \nu(dz) \right] \\
&= \mathbb{E}^* \left[-\beta T \int_{\mathbb{R}} \gamma(z) \exp(-\beta\gamma(z))\nu(dz) - T \int_{\mathbb{R}} \left\{ \exp(-\beta\gamma(z)) - 1 \right\} \nu(dz) \right] \\
&= T \int_{\mathbb{R}} 1 - \exp(-\beta\gamma(z))(1 + \beta\gamma(z))\, \nu(dz).
\end{aligned}$$

A numerical experiment is shown in Figure 1. Note that the expression $\exp(-\beta \dots)$ appears again in the expression for the optimal intensity measure. In this simple example explicit computations are possible. We aim to extend the theory developed in this paper to general stochastic processes in the near future.

4. The optimal control as a Malliavin derivative

Let $(\Omega, \mathscr{F}, \mathbb{Q})$ be a probability space on which a p-dimensional standard Brownian motion $(W_t)_{0 \le t \le T}$ is defined, with $T < \infty$. Let $(\mathscr{F}_t)_{t \ge 0}$ be the filtration generated by W. In this section we write $D_t F$ for the

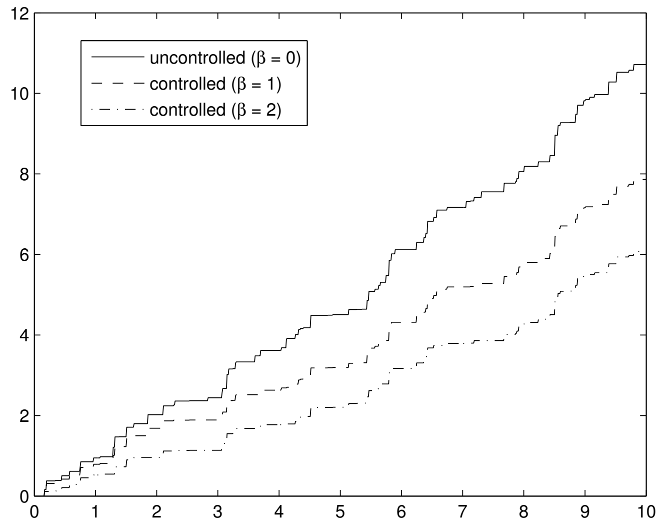

FIGURE 1. Illustration of the example of Section 3.4. Sample paths are shown of a controlled Lévy process ($\gamma(z) = z$) for different values of β. Here $\nu(dz) = f(z)\,dz$, with $f(z) = \alpha z^{-1/2}\exp(-\delta z)$. This results in $\nu^*(dz) = f^*(z)\,dz$ with $f^*(z) = \alpha z^{-1/2}\exp(-(\beta+\delta)z)$. Parameter values are $\alpha = 10$, $\delta = 5$. The sample paths for different values of β are constructed using the same underlying pseudorandomly generated numbers so that the sample paths may be compared. Note that for higher β, there are fewer jumps and the jump sizes are slightly smaller.

Malliavin derivative of an $\mathscr{F}_T$-measurable random variable F at time t and $\mathbb{D}^{1,q}$ for the domain of D in $L^q(\Omega)$, $q \geq 1$. See [DNØP09, Nua06] for details.

The following lemma is a consequence of [Nua06, Proposition 1.2.8].

Lemma 4.1. *Suppose $F \in \mathbb{D}^{1,2}$. Then $\mathbb{E}^{\mathbb{Q}}[F|\mathscr{F}_t] \in \mathbb{D}^{1,2}$ for $0 \leq t \leq T$ and $D_t\mathbb{E}^{\mathbb{Q}}[F|\mathscr{F}_t] = \mathbb{E}^{\mathbb{Q}}[D_tF|\mathscr{F}_t]$.*

Theorem 4.2. *Suppose $\mathbb{P}$ is absolutely continuous with respect to $\mathbb{Q}$ with Radon-Nikodým derivative $Z = \frac{d\mathbb{P}}{d\mathbb{Q}}$, where Z is $\mathscr{F}_T$-measurable for some $T > 0$. Let $Z_t := \mathbb{E}^{\mathbb{Q}}[Z|\mathscr{F}_t]$, $0 \leq t \leq T$, denote the density process. Suppose $Z \in \mathbb{D}^{1,2}$. Then $Z_t \in \mathbb{D}^{1,2}$ for all $0 \leq t \leq T$. Define a stochastic process V by*

$$V_t := D_t \ln Z_t = \frac{D_tZ_t}{Z_t} = \frac{\mathbb{E}^{\mathbb{Q}}[D_tZ|\mathscr{F}_t]}{\mathbb{E}^{\mathbb{Q}}[Z|\mathscr{F}_t]}, \quad 0 \leq t \leq T. \tag{11}$$

Then $Z = \exp\left(\int_0^T V_s\,dW_s - \frac{1}{2}\int_0^T V_s^2\,ds\right)$.

It is interesting to compare expression (11) to Proposition 3.8 (iii).

Proof. Note that

$$D_sZ_s = D_s\mathbb{E}^{\mathbb{Q}}[Z_t|\mathscr{F}_s] = \mathbb{E}^{\mathbb{Q}}[D_sZ_t|\mathscr{F}_s]\mathbb{1}_{[0,s]}(s) = \mathbb{E}^{\mathbb{Q}}[D_sZ_t|\mathscr{F}_s] \quad \text{for } 0 \leq s \leq t \in [0,T],$$

where the second equality is a consequence of Lemma 4.1. By the Clark-Ocone representation formula [Nua06, Proposition 1.3.14]

$$Z_t = 1 + \int_0^t \mathbb{E}^{\mathbb{Q}}[D_sZ_t|\mathscr{F}_s]\,dW_s = 1 + \int_0^t D_sZ_s\,dW_s = \int_0^T V_sZ_s\,dW_s$$

for $V_s = \frac{D_sZ_s}{Z_s}$. By the chain rule of Malliavin calculus, [Nua06, Proposition 1.2.3], $D_s \ln Z_s = \frac{D_sZ_s}{Z_s}$. Using Lemma 4.1 again, we have $\mathbb{E}^{\mathbb{Q}}[D_tZ|\mathscr{F}_t] = D_t\mathbb{E}^{\mathbb{Q}}[Z|\mathscr{F}_t] = D_tZ_t$ which finishes the argument. $\square$

An application of the chain rule of Malliavin calculus also gives the following corollary.

Corollary 4.3. *Suppose $C \in \mathbb{D}^{1,2}$ and let $\beta > 0$. Then the stochastic process U^* defined by*

$$U_t^* := -\frac{\beta \mathbb{E}^{\mathbb{Q}}\left[\exp(-\beta C) D_t C | \mathscr{F}_t\right]}{\mathbb{E}^{\mathbb{Q}}\left[\exp(-\beta C)|\mathscr{F}_t\right]}, \quad 0 \le t \le T,$$

solves Problem 3.1.

4.1. **Example:** $C = \frac{1}{2}(W_T - x^*)^2$**, continued.** We may apply Corollary 4.3 to the example of Section 3.3 as an alternative way to compute the optimal control process U^*. Using the chain rule of Malliavin calculus,

$$D_t C = \tfrac{1}{2} D_t (W_T - x^*)^2 = (W_T - x^*) D_t W_T = W_T - x^*, \quad 0 \le t \le T,$$

and it is then straightforward to check (using Lemma A.6 (ii)) that

$$-\beta \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\beta C\right) D_t C | \mathscr{F}_t\right] / \mathbb{E}^{\mathbb{Q}}\left[\exp(-\beta C)|\mathscr{F}_t\right]$$

gives the same expression for U_t^* as the one already obtained using Itô's formula.

4.2. **Example:** $C = \max_{0 \le t \le T} W_t$**.** We will now apply the results of Section 4 to an example where Itô's formula fails. This example therefore shows the strength of the new approach.

Define $M_t := \max_{0 \le s \le t} W_s$ and take $C := M_T$ for some $T > 0$. The optimal density function is $Z^* \propto \exp(-\beta C)$ and we wish to obtain the density process $\mathbb{E}^{\mathbb{Q}}[Z^*|\mathscr{F}_t]$. For the distribution of M_t we have (by [KS91, Section 2.8.A])

$$\mathbb{Q}(M_t \ge a) = \left(\frac{2}{\pi}\right)^{1/2} \int_{a/\sqrt{t}}^{\infty} \exp(-\xi^2/2)\, d\xi, \quad t \ge 0, a \ge 0.$$

Conditional on $\mathscr{F}_t$, the event $M_T = M_t$ occurs when the maximum over $[t, T]$ does not exceed $y := M_t$. This has the same probability as the event that the maximum over $[0, T-t]$ does not exceed $y - x$, for $x := W_t \le y$, so

$$\mathbb{Q}(M_T = M_t | W_t = x, M_t = y) = \mathbb{Q}(M_{T-t} \le y - x) = \left(\frac{2}{\pi}\right)^{1/2} \int_0^{\frac{M_t - W_t}{\sqrt{T-t}}} \exp(-\xi^2/2)\, d\xi.$$

For $0 \le x \le y < z$ we compute

$$\begin{aligned}\mathbb{Q}\left(M_T \ge z | W_t = x, M_t = y\right) &= \mathbb{Q}\left(M_T = M_t | W_t = x, M_t = y\right) + \mathbb{Q}\left(M_{T-t} \ge z - x\right) \\ &= \left(\frac{2}{\pi}\right)^{1/2} \int_0^{\frac{M_t - W_t}{\sqrt{T-t}}} \exp(-\xi^2/2)\, d\xi + \left(\frac{2}{\pi}\right)^{1/2} \int_{\frac{z - W_t}{\sqrt{T-t}}}^{\infty} \exp(-\xi^2/2)\, d\xi.\end{aligned}$$

Therefore the density function of M_T conditional on $\mathscr{F}_t$ is equal to

$$f_{M_T|\mathscr{F}_t}(\xi) = \left(\frac{2}{\pi(T-t)}\right)^{1/2} \exp\left(-\frac{(\xi - W_t)^2}{2(T-t)}\right), \quad \text{for } \xi > M_t \ge W_t.$$

Write $K := \mathbb{E}^{\mathbb{Q}}[\exp(-\beta M_T)]$. We will make use of the error function (erf) and complimentary error function (erfc), defined by

$$\operatorname{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x \exp(-\eta^2)\, d\eta, \quad \operatorname{erfc}(x) := 1 - \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} \exp(-\eta^2)\, d\eta, \quad x \ge 0.$$

We compute

$$\begin{aligned}Z_t^* &= \mathbb{E}^{\mathbb{Q}}[Z^*|\mathscr{F}_t] = \frac{1}{K}\mathbb{E}^{\mathbb{Q}}\left[\exp(-\beta M_T)|\mathscr{F}_t\right] = \frac{1}{K}\mathbb{E}^{\mathbb{Q}}\left[\exp(-\beta M_T)\mathbb{1}_{M_T = M_t}|\mathscr{F}_t\right] + \frac{1}{K}\mathbb{E}^{\mathbb{Q}}\left[\exp(-\beta M_T)\mathbb{1}_{M_T > M_t}|\mathscr{F}_t\right] \\ &= \frac{1}{K}\exp(-\beta M_t)\mathbb{Q}(M_T = M_t|\mathscr{F}_t) + \frac{1}{K}\mathbb{E}^{\mathbb{Q}}\left[\exp(-\beta M_T)\mathbb{1}_{M_T > M_t}|\mathscr{F}_t\right] \\ &= \frac{1}{K}\left[\exp(-\beta M_t)\left(\frac{2}{\pi}\right)^{1/2} \int_0^{\frac{M_t - W_t}{\sqrt{T-t}}} \exp(-\xi^2/2)\, d\xi + \left(\frac{2}{\pi(T-t)}\right)^{1/2} \int_{M_t}^{\infty} \exp(-\beta\xi) \exp\left(-\frac{(\xi - W_t)^2}{2(T-t)}\right) d\xi\right] \\ &= \frac{1}{K}\left[\exp(-\beta M_t) \operatorname{erf}\left(\frac{M_t - W_t}{\sqrt{2(T-t)}}\right) + \exp\left(-\beta W_t + \tfrac{1}{2}\beta^2(T-t)\right) \operatorname{erfc}\left(\frac{M_t - W_t + \beta(T-t)}{\sqrt{2(T-t)}}\right)\right]\end{aligned}$$

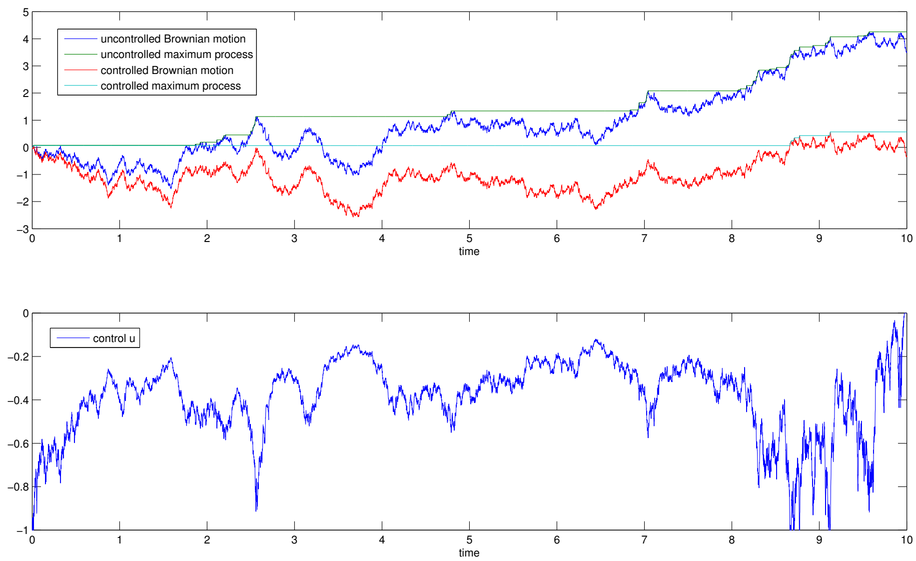

FIGURE 2. Above: sample path of a Brownian motion and its maximum process, uncontrolled and controlled ($\beta = 1$). Below: the control policy corresponding to the controlled Brownian motion.

An expression for the Malliavin derivative of M_T is available: $D_t M_T = \mathbb{1}_{[0,\tau]}(t)$, where τ is the a.s. unique point where W attains its maximum. See [Nua06, Exercise 1.2.11]. Hence by the chain rule for the Malliavin derivative,

$$D_t Z^* = -\frac{\beta}{K}\exp(-\beta M_T) D_t M_T = -\frac{\beta}{K}\exp(-\beta M_T)\mathbb{1}_{[0,\tau]}(t).$$

Note that $t < \tau$ if and only if $M_T > M_t$. We can compute

$$\begin{aligned}\phi_t := \mathbb{E}^{\mathbb{Q}}[D_t Z^* | \mathscr{F}_t] &= -\frac{\beta}{K}\mathbb{E}^{\mathbb{Q}}\left[\exp(-\beta M_T)\mathbb{1}_{\{M_T > M_t\}} \middle| \mathscr{F}_t\right] \\ &= -\frac{\beta}{K}\left(\frac{2}{\pi(T-t)}\right)^{1/2}\int_{M_t}^{\infty}\exp\left(-\beta\xi - \frac{(\xi - W_t)^2}{2(T-t)}\right) d\xi \\ &= -\frac{\beta}{K}\exp\left(-\beta W_t + \tfrac{1}{2}\beta^2(T-t)\right)\operatorname{erfc}\left(\frac{M_t - W_t + \beta(T-t)}{(2(T-t))^{1/2}}\right).\end{aligned}$$

By Theorem 4.2, we conclude that for $U_t^* := \phi_t / Z_t^*$ we have

$$dZ_t^* = U_t^* Z_t^* \, dW_t$$

so that U_t^* is the stochastic process which solves the optimization problem

$$\text{minimize} \quad J(U) := \mathbb{E}^U\left[\max_{0\le t\le T} W_t + \frac{1}{2\beta}\int_0^T u_s^2\right] = \mathbb{E}^U\left[\max_{0\le t\le T}\left(W_t^U + \int_0^t U_s \, ds\right) + \frac{1}{2\beta}\int_0^T U_s^2\right],$$

with W^U a Brownian motion under $\mathbb{P}^U$. Note that ϕ_t and Z_t^*, and hence U_t^*, are explicitly given in terms of t, W_t and M_t. The process U^* may be written 'succintly' as $U_t^* = u(t, W_t, M_t)$, where

$$u(t,w,m) = \frac{-\beta\exp\left(-\beta w + \frac{1}{2}\beta^2(T-t)\right)\operatorname{erfc}\left(\frac{m-w+\beta(T-t)}{\sqrt{2(T-t)}}\right)}{\exp(-\beta m)\operatorname{erf}\left(\frac{m-w}{\sqrt{2(T-t)}}\right) + \exp\left(-\beta w + \frac{1}{2}\beta^2(T-t)\right)\operatorname{erfc}\left(\frac{m-w+\beta(T-t)}{\sqrt{2(T-t)}}\right)}, \quad 0 \le t < T, w \in \mathbb{R}, m \ge w.$$

An illustration of this control policy and its effect on a sample path of W_t^U and M_t^U is provided in Figure 2.

Remark 4.4. This example illustrates how the theory applies to non-Markovian processes, and therefore provides a method that applies where a naive application of dynamic programming (i.e. the HJB equation) would fail.

In this particular case, by augmenting the state to (W_t, M_t) the optimal control becomes Markovian, but this requires a nonstandard application of the HJB equation; see also Remark 5.4 and [HS91] for a detailed analysis in a closely related class of examples.

5. Relation to classical stochastic optimal control

In this section we will link the theory of the previous sections to the classical theory of stochastic optimal control [FR75]. We will list some instances of optimal control problems and explain how the theory of the previous sections can be applied.

Assume as before the conditions of Hypothesis 3.2, defining a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t), \mathbb{Q})$ on which a p-dimensional standard Brownian motion W is defined. In classical stochastic optimal control theory the notion of state is fundamental. The dynamics of the state will be described by a stochastic differential equation. For this we require the following additional assumptions.

Hypothesis 5.1. Suppose $b : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^{n\times p}$ are

(i) *locally Lipschitz*, i.e. for every bounded set $B \subset \mathbb{R}^n$ and $T > 0$ there exists a constant $K > 0$ such that

$$|b(t,x) - b(t,y)| \leq K|x-y|, \quad \text{and} \quad ||\sigma(t,x) - \sigma(t,y)|| \leq K|x-y|, \quad \text{for all } 0 \leq t \leq T \text{ and } x,y \in B.$$

(ii) *monotone* in the following sense: for every $T > 0$ there exists a positive constant $K > 0$ such that for all $x \in \mathbb{R}^n$ and $t \in [0,T]$,

$$x^T b(t,x) + \tfrac{1}{2}||\sigma(t,x)||^2 \leq K(1+|x|^2).$$

Note that the monotonicity condition (ii) above is less restrictive than the linear growth condition which is more commonly found in the literature.

Under these assumptions, we will consider for $x \in \mathbb{R}^n$ the stochastic differential equation

$$\left\{ \begin{array}{ll} dX_t = b(t,X_t)\, dt + \sigma(t,X_t) dW_t, & t \geq 0, \\ X_0 = x. & \end{array} \right. \tag{12}$$

We may think of (12) as describing the *uncontrolled dynamics*. We make use of the following result on the existence of a unique strong solution to (12). See [Mao97, Theorem 2.3.6, Theorem 2.4.1].

Theorem 5.2. *Under the conditions of Hypothesis 5.1, for every $x \in \mathbb{R}^n$ there exists a unique solution (up to indistinguishability), denoted by $(X_t^x)_{t\geq 0}$, to* (12) *satisfying* $\sup_{0\leq t\leq T} \mathbb{E}^{\mathbb{Q}}|X_t^x|^2 \leq c_T(1+|x|^2)$ *for every $T > 0$ and some constant c_T depending on T.*

Such a process X^x is called a *Markov diffusion process.* Note that in general the Markov process is time inhomogeneous since the dynamics depends explicitly on time through b and σ. If b and σ do not depend explicitly on t then X^x is called a *time homogeneous Markov diffusion process.*

We consider the *set of Markov controls* $\mathscr{U}_{\mathrm{M}}$ which consists of mappings $u \in B([0,\infty)\times\mathbb{R}^n; \mathbb{R}^p)$ such that for all $x \in \mathbb{R}^n$ the stochastic process U^x defined by $U_s^x(\omega) := u(s, X_s^x(\omega))$ is in $\mathscr{U}$. For $u \in \mathscr{U}_{\mathrm{M}}$ we will write $\mathbb{P}^{x,u} := \mathbb{P}^{U^x}$, with U^x as above, and similarly $\mathbb{E}^{x,u} = \mathbb{E}^{U^x}$. Note that $\mathbb{P}^{x,u}$ depends on x through the definition of U_s^x.

For $u \in \mathscr{U}_{\mathrm{M}}$, the process $(X_t^x)_{t\geq 0}$ also satisfies the following SDE, of which we can think of as the *controlled dynamics*:

$$\left\{ \begin{array}{ll} dX_t = (b(t,X_t) + \sigma(t,X_t)u(t,X_t))\; dt + \sigma(t,X_t) dW_t^{x,u}, & t \geq 0, \\ X_0 = x. & \end{array} \right. \tag{13}$$

where by Girsanov's theorem [KS91, Corollary 3.5.2], as before, the process $(W_t^{x,u})$ defined by $W_t^{x,u} := W_t - \int_0^T u(s, X_s^x)\, ds$ is a standard Brownian motion with respect to the probability measure $\mathbb{P}^{x,u}$.

We will specialize to the situation where the cost random variable C also depends on the initial condition and is a functional of the paths of the stochastic process X^x for $x \in \mathbb{R}^n$. We will write $C^x(\omega) := c(X^x(\omega))$ where $c : C([0,\infty);\mathbb{R}^n) \to \mathbb{R}$ for $\omega \in \Omega$. The following examples of cost functionals are often used. Here ξ denotes any path in $C([0,\infty);\mathbb{R}^n)$.

(i) *Finite time horizon problem*: $c(\xi) = \int_0^T \phi(t, \xi(t))\, dt + \psi(\xi(T))$ for some $T > 0$, with $\phi \in B([0,T] \times \mathbb{R}^n;\mathbb{R})$, $\psi \in B(\mathbb{R}^n;\mathbb{R})$;
(ii) *Infinite time horizon problem with exit from a region*: $c(\xi) = \int_0^\tau \phi(t,\xi(t))\, dt + \psi(\tau, \xi(T))$, where $\tau = \inf\{t \geq 0 : \xi \notin G\}$ with $G \subset \mathbb{R}^n$ open, $\phi, \psi \in B([0,\infty) \times \mathbb{R}^n;\mathbb{R}^n)$;

The method outlined in the previous sections will enable us to find solutions to these problems, under certain conditions. These solutions will be compared to the solutions obtained by classical methods.

Define as before the *total cost function* by

$$J(x;u) := \mathbb{E}^{x,u} C^x + \frac{1}{\beta}\mathscr{H}(\mathbb{P}^{x,u};\mathbb{Q}) = \mathbb{E}^{x,u}\left[c(X^x) + \frac{1}{2\beta}\int_0^\infty |u(s,X_s^x)|^2\, ds\right], \tag{14}$$

for $x \in \mathbb{R}^n$ which is now specialized to Markov controls $u \in \mathscr{U}_{\mathrm{M}}$. Recall X^x satisfies (13) with $W^{x,u}$ a standard Brownian motion under $\mathbb{P}^{x,u}$. In this setting, we consider the following problem.

Problem 5.3 (Relative entropy weighted optimal control of Markov diffusion processes)**.** Find the value function $J^* : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}$, defined by

$$J^*(x) = \inf_{u \in \mathscr{U}_{\mathrm{M}}} J(x;u) \quad \text{for } t \in [0,\infty) \text{ and } x \in \mathbb{R}^n,$$

and, in case it exists, the optimal control policy $u^* \in \mathscr{U}_{\mathrm{M}}$ which satisfies

$$J^*(x) = J(x;u^*) \quad \text{for all } x \in \mathbb{R}^n.$$

It should now be clear that solving the relative entropy weighted Problem 5.3 is equivalent to solving a classical stochastic optimal control problem with quadratic control costs and with dynamics given by (13).

Remark 5.4. We cannot always expect to find Markov controls that are optimal in the sense of the more general Problem 3.1. For example, in Section 4.2, where a solution was computed for minimization of $M_T = \max_{0 \leq t \leq T} W_t$, the optimal stochastic process depends not only on time t and the current state W_t but also on the running maximum M_t. By augmenting the state to (W_t, M_t) the optimal control becomes Markovian. However, the time evolution of (W_t, M_t) cannot be put in the shape of (12). In fact, the process X defined by $X_t = M_t - W_t$, $0 \leq t \leq T$, has the same probability law as a Brownian motion reflected at the origin and satisfies a Skorohod equation; see [KS91, Section 3.6.C].

Remark 5.5. Often in the theory of stochastic optimal control, the cost function $J(x)$ is defined to depend explicitly on some starting time t_0. This explicit dependence on starting time is useful in obtaining the optimal control through the dynamic programming principle (i.e. the Hamilton-Jacobi-Bellman equation). Since we do not use the dynamic programming principle we do not need to consider the value function for all initial times t_0. In our setup we always have $t_0 = 0$ to avoid confusion, without loss of generality.

5.1. **Linearized Hamilton-Jacobi-Bellman equation.**

In this section sufficient conditions are obtained in order for the optimal control U^* to be a Markov policy, so that $U_t^* = u(t, X_t)$.

Let $G \subset \mathbb{R}^n$ be open and let $\tau^x := \inf\{t \geq 0 : X_t^x \notin G\}$ denote the exit time from G. Note that τ^x is a stopping time. We will study two cases: the infinite horizon problem and the time homogeneous exit problem.

5.1.1. *Finite horizon case.* Let $T < \infty$ and define the stopping time $\tau^x_T := \tau^x \wedge T$. Let $C^x = \int_0^{\tau^x_T} \phi(t, X^x_t)\, dt + \psi(\tau^x_T, X^x_{\tau^x_T})$ with $\phi, \psi \in B_b([0,\infty)\times\mathbb{R}^n)$ so that the integral exists and is finite. Let L denote the *Kolmogorov backward operator* corresponding to (12), given by

$$Lf(t,x) = \frac{\partial f}{\partial t}(t,x) + \sum_{i=1}^n b^i(t,x)\frac{\partial f}{\partial x^i}(t,x) + \tfrac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \left[\sigma(t,x)\sigma(t,x)^T\right]^{ij} \frac{\partial^2 f}{\partial x^i x^j}(t,x) \tag{15}$$

for $f \in C^{1,2}([0,T]\times\mathbb{R}^n;\mathbb{R})$.

Theorem 5.6 (Linearized Hamilton-Jacobi-Bellman equation – Finite horizon)**.** *Let Hypothesis 5.1 be satisfied. Suppose that $y \in C^{1,2}([0,T]\times\mathbb{R}^n;\mathbb{R})$ satisfies the PDE*

$$\begin{cases} Ly(t,x) - \beta\phi(t,x)y(t,x) = 0, & t \in [0,T], x \in G, \\ y(t,x) = \exp(-\beta\psi(t,x)), & x \in \partial G \text{ or } t = T. \end{cases} \tag{16}$$

Suppose, $\mathbb{Q}$-almost surely, $y(t, X^x_t) > 0$ for $0 \le t \le \tau^x_T$ and $x \in G$. Define $u^*(t,x) := \frac{\sigma(t,x)^T \nabla y(t,x)}{y(t,x)}$ for all $(t,x) \in ([0,T],G)$ for which $y(t,x) \neq 0$. Then any measurable extension of u^* to $[0,T]\times G$ solves Problem 5.3. Furthermore the value function of Problem 5.3 is given by $J^*(x) = -\frac{1}{\beta}\ln y(0,x)$, for $x \in G$.

Proof. Fix $x \in \mathbb{R}^n$ and omit the 'x'-superscripts in X^x_t, C^x etc and the '$*$'-superscript in u^*. Define $Y_t := y(t, X_t)$. Then by Itô's formula,

$$\begin{aligned} dY_t &= Ly(t,X_t)\, dt + \sum_{i=1}^n\sum_{j=1}^p \sigma^{ij}(t,X_t)\frac{\partial y}{\partial x^i}(t,X_t)\, dW^j_t \\ &= \beta\phi(t,X_t)y(t,X_t)\, dt + y(t,X_t)\sum_{j=1}^p u^j(t,X_t)\, dW^j_t \\ &= \beta\phi(t,X_t)Y_t\, dt + Y_t\sum_{j=1}^p u^j(t,X_t)\, dW^j_t. \end{aligned}$$

It may be checked that a solution for this SDE is given by

$$y(t,X_t) = Y_t = Y_0\exp\left(\beta\int_0^t \phi(s,X_s)\, ds + \sum_{j=1}^p\int_0^t u^j(s,X_s)\, dW^j_s - \tfrac{1}{2}\int_0^t |u(s,X_s)|^2\, ds\right).$$

By the boundary condition,

$$Y_{\tau_T} = \exp(-\beta\psi(\tau_T, X_{\tau_T})) = Y_0\exp\left(\beta\int_0^{\tau_T}\phi(s,X_s)\, ds + \sum_{j=1}^p\int_0^{\tau_T} u^j(s,X_s)\, dW^j_s - \tfrac{1}{2}\int_0^{\tau_T}|u(s,X_s)|^2\, ds\right).$$

Multiplying by $\exp\left(-\beta\int_0^{\tau_T}\phi(s,X_s)\, ds\right)$ yields

$$Y_0\exp\left(\sum_{j=1}^p\int_0^{\tau_T} u^j(s,X_s)\, dW^j_s - \tfrac{1}{2}\int_0^{\tau_T}|u(s,X_s)|^2\, ds\right) = \exp\left(-\beta\int_0^{\tau_T}\phi(s,X_s)\, ds - \beta\psi(\tau_T, X_{\tau_T})\right).$$

By taking expectations, $y(0,x) = Y_0 = \mathbb{E}^{\mathbb{Q}}\exp\left(-\beta\int_0^{\tau_T}\phi(s,X_s)\, dt - \beta\psi(\tau_T, X_{\tau_T})\right) = \mathbb{E}^{\mathbb{Q}}\exp(-\beta C)$. By the proof of Theorem 3.3 we may conclude that $U_t := u(t,X_t)$ solves Problem 3.1 and therefore that u solves Problem 5.3. □

Remark 5.7. Assume for simplicity that $a := \sigma\sigma^T$ and b are globally Lipschitz continuous and bounded, ϕ is bounded uniformly Hölder continuous and ψ is bounded continuous on $([0,T]\times\delta G)\cup(T\times\overline{G})$. Note that under the following further assumptions (16) has a unique solution that may be represented by the Feynman-Kac formula,

$$y(t,x) = \mathbb{E}^{t,x}\left[\exp\left(-\beta\int_t^{\tau^x_T\vee t}\phi(s,X_s)\, ds - \beta\psi(\tau^x_T\vee t, X_{\tau^x_T\vee t})\right)\right], \tag{17}$$

(where $\mathbb{E}^{t,x}$ denotes expectation with respect to the law of X satisfying (12) with initial condition $X_t = x$):

(i) nondegenerate case: G is open and bounded with smooth boundary, or $G = \mathbb{R}^n$, $a = \sigma\sigma^T$ is uniformly elliptic, i.e. $z^T a(t,x) z \geq \mu |z|^2$ for some $\mu > 0$ and all $z \in \mathbb{R}^n$, $t \in [0,T]$ and $x \in \overline{G}$ (see [Mao97, Theorems 2.8.2 and 2.8.3]); or

(ii) degenerate case: $G = \mathbb{R}^n$, b, σ, ϕ and ψ are homogeneous in t, in which case the Markov semigroup corresponding to X is a Feller-Dynkin semigroup corresponding to X (see [Kal02, Theorems 21.11 and 24.11]).

In case (17) holds, it follows immediately (by boundedness of ϕ and ψ) that $y > 0$ on its domain.

Remark 5.8. The expression (16) may be seen as a linearized version of the Hamilton-Jacobi-Bellman (HJB) equation. This linearized HJB equation may alternatively be obtained from the nonlinear HJB equation [FR75, Theorem 4.1], by applying a logarithmic transform (see [Kap05]). This observation formed the starting point for the research of this paper.

Remark 5.9. Note that if we do not restrict ourselves to Markov controls, the value function is given by (4), i.e. $J^*(x) = -\frac{1}{\beta} \ln \mathbb{E}^{\mathbb{Q}} \exp(-\beta C^x)$, which may be compared to (17). Theorem 5.6 is therefore partly a restatement of the Feynman-Kac solution to the corresponding Dirichlet and Cauchy problems [KS91, Section 5.7] (and the proof is largely similar). But note that the theorem also states the existence of a Markov control and gives an expression for the optimal Markov control u^* in this case.

5.1.2. *Exit problem.* We will now consider the time homogeneous case where b and σ do not depend on t, with G a bounded open subset of $\mathbb{R}^n$ with smooth boundary ∂G. Let L denote the time homogeneous Kolmogorov backward operator corresponding to (12), given by

$$Lf(x) = \sum_{i=1}^n b^i(x) \frac{\partial f}{\partial x^i}(x) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \left[\sigma(x)\sigma(x)^T\right]^{ij} \frac{\partial^2 f}{\partial x^i \partial x^j}(x) \tag{18}$$

for $f \in C^2(\mathbb{R}^n; \mathbb{R})$. Define the cost random variable by $C^x := \int_0^{\tau^x} \phi(X_s^x)\, ds + \psi\left(X^x_{\tau^x}\right)$.

Analogously to Theorem5.6 the following result can be established.

Theorem 5.10. *Let Hypothesis 5.1 be satisfied. Suppose that $y \in C^2(G;\mathbb{R})$ satisfies the PDE*

$$\begin{cases} Ly(x) - \beta\phi(x) y(x) = 0, & x \in G, \\ y(x) = \exp(-\beta\psi(x)), & x \in \partial G. \end{cases} \tag{19}$$

Suppose furthermore that $y(x) > 0$ for $x \in \overline{G}$. Define $u^*(x) := \frac{\sigma(x)^T \nabla y(x)}{y(x)}$ for all $x \in G$. Then u^* solves Problem 5.3. Furthermore the value function of Problem 5.3 is given by $J^*(x) = -\frac{1}{\beta} \ln y(x)$, for $x \in G$. In particular, y may be expressed as

$$y(x) = \mathbb{E}^{\mathbb{Q}}\left[\exp(-\beta C^x)\right] = \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\beta \int_0^{\tau^x} \phi(X_s^x)\, ds - \beta\psi(X_\tau^x)\right)\right]. \tag{20}$$

Remark 5.11. For example in the case of bounded G, with uniform ellipticity of $a := \sigma\sigma^T$, with a and b globally Lipschitz, ϕ bounded uniformly Hölder continuous and nonnegative, and ψ continuous on ∂G, by [Mao97, Theorem 2.8.1] a unique solution to (19) exists and is hence given by (20) guaranteeing positivity of y on $\overline{G}$.

Appendix A. Appendix

A.1. **Relative entropy.** In the following, μ and ν denote probability measures over some measurable space (S, Σ).

Definition A.1. Suppose, for all $A \in \Sigma$, that $\mu(A) = 0 \Rightarrow \nu(A) = 0$. We then say that ν is *absolutely continuous* with respect to μ. This is denoted as $\nu \ll \mu$. If both $\mu \ll \nu$ and $\nu \ll \mu$ then we say that μ and ν are *equivalent*.

We recall the classical notion of Radon-Nikodým derivative [Wil91, Theorem 14.13].

Theorem A.2 (Radon-Nikodým derivative)**.** *If $\nu \ll \mu$ then there exists a random variable $X \in \mathscr{L}^1(S, \Sigma, \mu)$ such that*

$$\nu(A) = \int_A X \, d\mu.$$

This variable X is called a version of the Radon-Nikodým derivative *of ν relative to μ on (S, Σ), and different versions agree μ-almost surely. We write*

$$\frac{d\nu}{d\mu} := X \quad \text{on } \Sigma, \quad \mu\text{-a.s.}$$

Definition A.3. We call a Σ-measurable nonnegative random variable f a *density (function)* with respect to μ if there exists a probability measure ν that has Radon-Nikodým derivative f relative to μ.

Definition A.4. The *relative entropy* of μ with respect to ν is defined as

$$\mathscr{H}(\mu; \nu) := \begin{cases} \int_S \ln\left(\frac{d\mu}{d\nu}\right) d\mu, & \text{if } \mu \text{ is absolutely continuous with respect to } \nu, \\ \infty, & \text{otherwise.} \end{cases} \tag{21}$$

Relative entropy is also known as *Kullback-Leibler divergence*; for this paper we use the term 'relative entropy' since it seems to be better known in the mathematics community. In general, $\mathscr{H}(\mu; \nu)$ is not symmetric in μ and ν.

The following proposition summarizes some useful properties of relative entropy. In particular, it indicates that relative entropy is a good indication of how similar two probability measures are.

Proposition A.5.

(i) *The relative entropy $\mathscr{H}(\mu; \nu)$ is well-defined (i.e. the integrals exist) and assumes its values within* $[0, \infty]$.

(ii) $\mathscr{H}(\mu; \nu) = 0$ *if and only if $\mu = \nu$ on S, μ-almost everywhere.*

(iii) $\mathscr{H}(\mu; \nu)$ *is strictly convex in μ.*

Proof. See [DE97, Section 1.4]. □

A.2. **Exponents of Gaussian random variables.** For convenience we state, without proof, the following lemma. (Note that the second identity is obtained by differentiating the first identity with respect to α and then setting $\alpha = 0$.)

Lemma A.6. *Suppose Y is a real-valued random variable that is normally distributed with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$. Let $\alpha \in \mathbb{R}$ and $\gamma \in \mathbb{R}$ such that $\rho := 1 + \gamma\sigma^2 > 0$. Then*

(i) $\mathbb{E}\left[\exp\left(\alpha Y - \frac{1}{2}\gamma Y^2\right)\right] = \frac{1}{\rho^{1/2}} \exp\left(-\frac{\gamma\mu^2 - 2\alpha\mu - \alpha^2\sigma^2}{2\rho}\right)$,

(ii) $\mathbb{E}\left[Y \exp\left(-\frac{1}{2}\gamma Y^2\right)\right] = \frac{\mu}{\rho^{3/2}} \exp\left(-\frac{\gamma\mu^2}{2\rho}\right)$.

Acknowledgements. We are grateful to dr. ir. O. van Gaans (Mathematical Institute, Leiden University) for his review of our work and his helpful suggestions that enabled us to improve the exposition.

References

[BD98] Michelle Boué and Paul Dupuis. A variational representation for certain functionals of Brownian motion. *Ann. Probab.*, 26(4):1641–1659, 1998.

[Dav79] M. H. A. Davis. Martingale methods in stochastic control. In *Stochastic control theory and stochastic differential systems (Proc. Workshop, Deutsch. Forschungsgemeinsch., Univ. Bonn, Bad Honnef, 1979)*, volume 16 of *Lecture Notes in Control and Information Sci.*, pages 85–117. Springer, Berlin, 1979.

[DE97] Paul Dupuis and Richard S. Ellis. *A weak convergence approach to the theory of large deviations.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, 1997. A Wiley-Interscience Publication.

[DNØP09] Giulia Di Nunno, Bernt Øksendal, and Frank Proske. *Malliavin calculus for Lévy processes with applications to finance.* Universitext. Springer-Verlag, Berlin, 2009.

[Fle82] Wendell H. Fleming. Logarithmic transformations and stochastic control. In *Advances in filtering and optimal stochastic control (Cocoyoc, 1982)*, volume 42 of *Lecture Notes in Control and Inform. Sci.*, pages 131–141. Springer, Berlin, 1982.

[FR75] Wendell H. Fleming and Raymond W. Rishel. *Deterministic and stochastic optimal control*. Springer-Verlag, Berlin, 1975. Applications of Mathematics, No. 1.

[FS06] Wendell H. Fleming and H. Mete Soner. *Controlled Markov processes and viscosity solutions*, volume 25 of *Stochastic Modelling and Applied Probability*. Springer, New York, second edition, 2006.

[Hau86] U. G. Haussmann. *A stochastic maximum principle for optimal control of diffusions*, volume 151 of *Pitman Research Notes in Mathematics Series*. Longman Scientific & Technical, Harlow, 1986.

[HS91] Arthur C. Heinricher and Richard H. Stockbridge. Optimal control of the running max. *SIAM J. Control Optim.*, 29(4):936–953, 1991.

[Kal02] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[Kap05] H.J. Kappen. Linear theory for control of nonlinear stochastic systems. *Physical review letters*, 95(20):200201, 2005.

[KS91] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.

[Mao97] X. Mao. Stochastic Differential Equations and Applications. *Horwood, Chichester*, 1997.

[Nua06] D. Nualart. *The Malliavin Calculus And Related Topics*. Springer, second edition, 2006.

[ØS07] B. Øksendal and A. Sulem. *Applied stochastic control of jump diffusions*. Universitext. Springer, Berlin, second edition, 2007.

[Tod06] Emanuel Todorov. Linearly-solvable markov decision problems. In *NIPS*, pages 1369–1376, 2006.

[Wil91] D. Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.